Measuring the Orthographic Vocabulary Size and Development of Japanese University Students

Craig JONES

Introduction

Vocabulary is arguably the most important aspect of learning a foreign language. In order to communicate in a foreign language a learner must have access to thousands of words. These words are the bricks that build communicative competence. Of course, bricks need cement to hold them together, and this is the function of grammar, but it is vocabulary that provides a learner's output with decoration, colour and even power. It frees a speaker to be creative and expressive in subtle ways. 'Without grammar very little can be conveyed, without vocabulary nothing can be conveyed' (Wilkins, 1972, p. 111).

Knowing the vocabulary size of a student or class can be extremely beneficial for a teacher and yet it is often overlooked in ESL programmes. Having an idea of how many words a student knows can allow educators to issue appropriate class material. Presenting learners with inappropriate vocabulary can have a negative washback effect on the class: vocabulary that is too difficult can be demoralising, and vocabulary that is too easy can demotivate students. It is essential to keep students stimulated for optimal language acquisition. Having a reliable method of measuring vocabulary growth is beneficial for educators and satisfying for students.

This study focused on the vocabulary size of university students in Japan. It uses a vocabulary size test called X-Lex (Meara & Milton, 2003). This particular test was selected as it has proven effective in estimating vocabulary size. Moreover, it is teacher and student friendly as it is quick and efficient to administer and mark.

It randomly selects 20 words from the first five 1000 word frequency bands. It is straightforward to develop multiple versions of the test, making it possible to measure students' vocabulary development over time. David (2008) found a strong correlation between checklist tests using different words from the same frequency bands.

The present study measured the vocabulary of first and second grade university students in May and October 2020: all participants were non-English majors. The study provided cross-sectional data to investigate how vocabulary develops over two years of university. The study also investigated correlations between the X-Lex (Meara & Milton, 2003) test and results achieved on TOEIC and their English university entrance exam.

Literature review

The size of a person's vocabulary relates to other aspects of language proficiency in both L1 and L2. There have been a large quantity of studies demonstrating that vocabulary size correlates with reading ability (Arnauld, 1992; Hu & Nation, 1992; Aizawa & Iso, 2013; Koda, 1989; Laufer, 1992; Mochizuki, 2010). Staer, (2008) found that as well as relating to reading there was also strong correlation between vocabulary size and knowledge of writing and grammar, with a moderate correlation to listening ability. There have also been studies conducted to investigate the relationship between vocabulary size and scores on well established international tests of English proficiency. (Chiang, 2018; Kanzaki, 2010; Mizumoto & Shimamoto, 2008; Taguchi, 2015) Discovered that vocabulary size correlated with TOEIC scores. It is evident that having a large vocabulary is beneficial when learning a language.

It is clear that being able to measure a person's vocabulary size is extremely beneficial for educators, researchers and learners. However, understanding a person's word knowledge is more complex than it would seem. In recent times L2 vocabulary research has been making distinctions between depth and breadth of

word knowledge: this idea was suggested by Anderson and Freebody (1981). Breadth of word knowledge is the number of words that a person knows; depth of knowledge refers to what the person knows about these words. A learner may hear a word in a film and understand the meaning but would be unable to produce the same word in a real situation. In such a scenario could it be said that the person truly knows the word? Nation (2001) states that a proficient learner understands a word's peripheral meanings such as pronunciation, inflections, parts of speech etc.

Nation (2001) classifies word knowledge in three categories: *form, meaning* and *use*: there are also three sub categories to each of these. Under the category *form* we have spoken and written knowledge of the word and word parts. Under the category *meaning* we have form and meaning; concepts; referents and associations. Under *use* we have grammatical functions, collocations and constraints (Nation, 2001, p. 22). It can be difficult to produce an all encompassing test to record this kind of Knowledge. Nation, acknowledges that multiple separate tests would be required to gain a comprehensive understanding of a particular learner's vocabulary knowledge (Nation, 2007, p. 39). Anderson and Freebody (1981) stated that a persons' vocabulary size can be considered as the number of words that the person knows at least some of the significant aspects of meaning. Nation (2001) also suggests that if a learner can recognise the form of a word and can attach meaning to it then it is part of that person's lexicon.

In recent times there has been a shift in vocabulary research toward focussing on vocabulary size. Meara & Wolter (2004) bemoan the focus on breadth and depth. They believe that there is too much focus on individual words and it is better to focus on lexicon has a whole. They believe that vocabulary size is the only dimension of any real importance, particularly when dealing with a smaller lexicon (Meara & Wolter, 2004, p. 45). Read (2000) concurs that vocabulary size gives a more in-depth idea of the overall state of the vocabulary, which is more useful than focusing on a limited number of words. This is apt for the present study as the participants are non-English majors likely to have a lower vocabulary size than

advanced learners.

The idea of measuring vocabulary size as a whole is all well and good, but how do we know what words should be counted? For example, an early estimate of vocabulary size declared that a native speaker of English knows 200000 words (Seashore & Eckerson, 1940). This number is daunting for any learner about to embark on an L2 journey. The reason that the number is so high is because such numbers were arrived at by counting all the words in a dictionary: every form of the word was counted as a different word, so the word *know, known* and *knows* would count as three separate words (Milton, 2009, p. 7). Modern vocabulary research focuses on word families, where all three words in the above example would be counted as one word. Vocabulary tests such as Nation's Levels Test (Nation, 1990) and the test used in this study, X-Lex (Meara & Milton, 2003) use lemmatisation in their word counts. A Lemma includes a head word and its most frequent inflections. However, the part of speech from the headword must not change. For example, the lemma of the verb *govern* would include *governs* and *governing*: these all count as one word, but *government* is a noun, therefore it would count as a different word (Milton, 2009, p. 10). Vermeer (2004) believes that the lemma is the most reliable method for counting words. There are some tests that use word families which are not concerned with how the part of speech changes, so in the above example *government* would be counted as the same word. It is important that researchers and educators are familiar with how words are counted when they use a particular test.

It is beneficial for researchers and educators to know what vocabulary size is needed to be proficient in English. The common thinking in vocabulary research is that there is a link between word frequency and learning. The idea is that the more frequent a word is in the target language, the earlier it will be acquired. Therefore, it has been said that the most frequent 2000 words in English are the most useful for a learner as they account for around 80% of any normal text (Milton, 2009, p. 47). This knowledge prompted Nation to suggest that everything possible should be done to acquire these words in the early part of a learners development (Nation,

2001, p. 16).

This amount of coverage is an achievable target for a beginner learner. Laufer (1989) calculated that a learner needs to know 95% of the words in a text to achieve 'reasonable' comprehension. Hu and Nation (2000) suggested that to read for pleasure 98% coverage is needed. Laufer (1989) investigated how many lemmatised words a learner would need in order to achieve 95% coverage. She concluded that knowing 4800 lemmatised words would give a learner 98% coverage of most texts. However, there are many different kinds of texts. For example, Aizawa & Iso (2008) investigated how much vocabulary was needed to study academic texts. They found that 6500 words were needed to achieve a score of 80% on the TOEFL reading comprehension test. These numbers can be debated, and as mentioned earlier: what constitutes a word and how the researcher is counting can vary. However, such estimates provide a useful number for educators, researchers and learners to use as a target.

After establishing the importance of vocabulary size we need to look at the most consistent and accurate methods for measuring it. The most widely used ESL test is the Vocabulary Levels Test (Nation, 1990). It was originally designed as a diagnostic test for educators, but is now also widely used by researchers. Schmitt, et al (2001) updated the levels test making it shorter and more practical. The test is multiple choice containing 30 items in ten clusters at five different frequency bands. Rather than giving an overall vocabulary size it provides frequency information. In Japan numerous orthographic tests have been tried and tested (Aizawa, 1998; Mochizuki, 1998; Sato, 2003). A multitude of vocabulary research has been done in Japan using these tests. The above tests involve reading a word and linking it to its meaning either in the L1 or L2. Another style of test that is commonly used in vocabulary research is the checklist or yes/no test. The Meara & Milton (2003) test used in this study is a checklist test.

The format of a checklist test is simple. The participant is presented with a word and is then asked to check a box declaring whether they know the word or not.

This kind of testing is useful for educators and researchers as it is easy to create, distribute and complete. This makes it easier to test a large number of words. It also becomes easy to make multiple versions of the test, because words are selected randomly from the first five 1000 word frequency bands. Some examples of checklist tests are the Eurocentre's Vocabulary Size Test (Meara & Jones, 1990) and the X-Lex, Swansea University Levels Test (Meara & Milton, 2003).

The present study recorded cross-sections of Japanese university students' vocabulary size over a two year period. Before analysing the data it is important to know what kind of vocabulary development is usual. In any language course regular input of vocabulary can be expected (Milton, 2009). Milton and Meara (1998) reviewed research papers from Japan and Europe reporting on vocabulary growth. They found that learners' vocabulary size usually grew at a rate of 500 to 600 words in each year of formal study. Milton (2006) conducted a cross-sectional study of 227 students at a Greek language school. He found that over a seven year period the vocabulary size grew by 500 words each year. Orosz (2009) studied Hungarian learners finding that vocabulary size increased by 300 to 400 words. Sato (2017) found consistent growth in Japanese junior high school students over three years. Vocabulary growth can vary as it does not take into account factors like age and contact hours per week. Meara and Milton (1998) considered such factors and found that if vocabulary uptake was analysed per classroom hour a clear and consistent number emerges. The number that they arrived at was four words per contact hour. Milton (2009) reflected on this number and found that after reviewing several of the studies mentioned above 'the figure of about four words learned per contact hour seems like an extremely useful yardstick for teachers and learners' (Milton, 2009, p. 88). It is important to note that this is the rate that vocabulary is acquired not introduced. Gairns & Redman (1986) suggest that 8 to 12 words per hour would be a reasonable target for learners.

Research aims

1. To measure the vocabulary size of first and second grade university students.
2. To assess whether the X-Lex (Meara & Milton, 2003) vocabulary test is suitable for measuring the vocabulary size of Japanese university students.
3. To measure whether the students' vocabulary size develops consistently over two years.
4. To investigate whether the X-Lex test scores correlate with other examinations of English proficiency.

Methodology and participants

The first goal of the study was to measure the vocabulary size of the first and second grade university students. A total of 402 participants from a public university in central Japan participated in the study. The participants were all first and second grade non-English majors. In May 2020, 113 grade one students and 101 grade two students completed the X-Lex (Meara & Milton, 2003) vocabulary test. At the start of the second semester in October 2020 a further 102 grade one students and 86 grade two students took the test (the start of both semesters was delayed by one month due to the Covid-19 pandemic).

The participants were also asked to provide their English score from the university entrance exam. The second year participants were also asked to provide their most recent TOEIC score. The students were under no obligation to provide these scores, and many participants completed the X-Lex test but did not offer their TOEIC and entrance test results. A total of 139 participants provided their English entrance exam results, and 79 second year participants provided TOEIC scores.

Research instrument

The vocabulary size test used in the study was the X-Lex (Meara & Milton, 2003). The version of the test used in this study was taken from Milton (2009). The test presented the participants with 120 words. They were asked whether they knew

the word or not. If they knew the word they were instructed to check yes, and if they did not they were to check no. The X-Lex test randomly selects 20 words from each of the first five 1000 word frequency bands; this takes the item count to 100. Then 20 pseudo-words are included: the words are designed to sound like English words but are not. Some example pseudo-words are *frequid, horobin* and *cantileen* (Milton, 2009, p. 254). The pseudo-words are included in the test to counter guess work. The participants were awarded 50 points for every yes box they checked giving a raw score out of 5000. Participants were deducted 250 points for every pseudo-word that they checked yes. The initial plan for this study was for the participants to sit a paper version of the test. However, due to the Covid-19 pandemic all classes were moved online. Therefore, the participants were given an online version of the test that they could submit from home. A checklist test always relies on the participants being honest. The participants completed this test voluntarily so there is no reason to suspect that they did not follow the test rules correctly.

Results and discussion

Vocabulary Size

Table 1: Descriptive statistics for vocabulary size

Group	n	Mean	Standard Deviation
1st Grade (May)	113	4174.34	414.853
2nd Grade (May)	101	3949.01	407.430
1st Grade (October)	102	4023.53	448.137
2nd Grade (October)	86	3980.81	396.838

The descriptive statistics of the X-Lex (Meara & Milton, 2003) vocabulary size test are presented in Table 1. The average vocabulary size of the grade one students at the beginning of university was 4174. The number is lower than Laufer's (1989) 4800, but it gives a very good base for development. The university where

this study was conducted requires non-English majors to take two 90 minute English classes a week. If we use the research mentioned earlier by Meara and Milton (1998) and set a rather conservative estimate of four words acquired per teaching hour we could expect the vocabulary size to increase by 360 in the first year. Using this model by the end of the second year of university Laufer's (1989) 98 percent coverage becomes an achievable target. However, by looking at the mean vocabulary size in Table 1 we can see that the vocabulary size does not grow at the expected rate.

Table 1 shows that the vocabulary size decreases over the first two years. The vocabulary size is 4174 for grade one students in their first week of university, 4024 at the start of the second semester in October, 3980 at the start of the second grade in October. The data shows that rather than an expected increase of around 360 words we have a decrease of 225 words. To test whether this decrease in vocabulary size was significant a t-test was performed. The assumption of homogeneity of variance was tested and satisfied via Levene's F test, $F(197) = .18$, $p = .665$. The independent samples t-test was associated with a statistically significant effect, $t(34) = 3.32$, $p = .001$. Thus, it can be said that the vocabulary size was significantly lower at the start of the second year than it was at the start of the first year.

There are several possible reasons for the decline. Students in Japan begin studying English in grade one of elementary school and continue until university. The vocabulary issued is recommended by the government. There are also factors like external English examinations such as EIKEN to consider. For example, passing a certain level of EIKEN can be beneficial when applying for certain high schools. The vocabulary required to pass the various EIKEN levels increases in number and difficultly at each stage. High school students applying for university are also expected to sit an English exam, even if they are not intending to study English. The high stakes of entrance exams compels students to focus on studying vocabulary. This could explain why students start university with such a large vocabulary.

At university the incentive to study vocabulary is not as powerful. First of all the students in the current study were non-English majors. In some cases they were majoring in a different language like French or Spanish. This means they have to focus on learning vocabulary in a new language. All non-English majors are expected to study English for three hours a week divided into two 90 minute classes. The first class focuses on reading and writing and the second class focuses on communication. The teacher is different for each class and students rotate every fifteen weeks. This is done so that the students get to experience as many different teaching styles as possible in their two years of English study. Despite sharing course goals and outcomes, there is no coordination when it comes to vocabulary.

The students may not be as motivated to learn and retain vocabulary as they were in high school but educators can increase motivation by incentivising vocabulary study. The first thing that teachers or course developers need to do is set clear goals for how much vocabulary is to be introduced; just as it is with other aspects of language learning. It is possible for teachers and course developers to incentivise the study of vocabulary. This can be as simple as creating a vocabulary test to complete at the end of the semester and assigning it a percentage of the final grade. This would likely compel students to focus on vocabulary retention.

The Vocabulary Test

Table 2: Number of correct answers in each frequency band by grade one participants in May

	1000	2000	3000	4000	5000
Correct answers (total 2340)	2325 (99.3%)	2119 (90.5%)	2105 (89.9%)	1862 (79.9%)	1487 (63.5%)

The X-Lex (Meara & Milton, 2003) test focuses on the most frequent 5000 words in English. The frequency model suggests that the more frequent a word the earlier it will be encountered and learned (Milton, 2009, p. 25). Wesche & Paribakht

(1996) suggest that this is an old idea assumed by researchers and teachers that has only been proven in more recent times. Meara (1992) developed a frequency profile capable of supporting the theory.

A frequency profile for the grade one participants was created to investigate whether the data from the X-Lex (Meara & Milton, 2003) supported the theory. The correct answers in each 1000 word frequency band were counted and are presented in Table 2 above. The data shows that the vocabulary knowledge gets lower in each frequency band which supports the frequency model. A number of studies have been conducted demonstrating that the theory is strong (Aizawa, 2006; Milton, 2006; Richards & Malvern, 2007).

Figure 1 presents the frequency profile of the grade one participants in the present study. The first column represents knowledge of the first thousand most frequent words, the second column represents knowledge of the second thousand most frequent words, and so on. The theory suggests that a learner should know more of the high frequency words than the low frequency words. Therefore, the columns in a frequency profile should start higher on the left and get gradually lower as they move to the right.

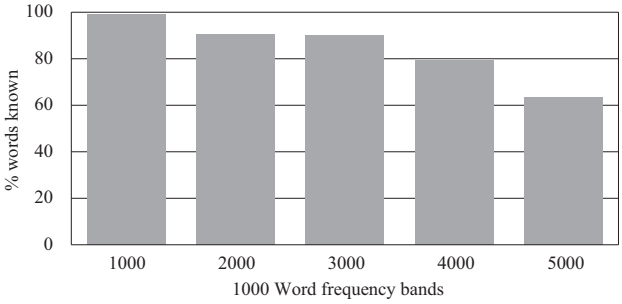

Figure 1: Frequency profile for grade one participants in May

As a learner's vocabulary increases the columns on the right of the frequency profile should increase until they reach a ceiling effect of 100%. A look at Figure 1 shows that the vocabulary knowledge of the first thousand words was 99.3%, the

second thousand was 90.5% and the third thousand was 89.9%. These numbers are extremely high and come close to achieving the ceiling effect. Furthermore, the fourth column at 79.9% and the fifth column at 63.5% are also relatively high. This suggests that the test doesn't represent the vocabulary size of the participants, but it does give an idea of their knowledge of the first 5000 words. That does not detract from the test, as this is what it is designed to do. However, it does appear that the same test would be more effective for these participants if it focused on higher frequency bands. Perhaps starting with the 3000 word frequency band and ending with the 7000 word frequency band. The maximum score of 5000 would appear to be a limitation for students of this level. Aizawa (2006) used the JACET 8000 word list to create a test of the first eight 1000 word frequency bands with a maximum score of 8000. The higher score would appear to be more appropriate for more advanced learners. However, a larger test may take more time to create and administer. Therefore, a teacher or researcher looking to use an expanded test would have to take this into consideration.

Correlation between vocabulary size and the university English entrance exam

In order to measure the effectiveness of the X-Lex (Meara & Milton, 2003) a comparison was made to other measures of English proficiency. Table 3 displays descriptive statistics for the grade one participants that provided their English entrance exam results. Table 4 presents the descriptive statistics for the grade two participants that provided TOEIC scores. A Pearson correlation test was conducted to compare both sets of results with the vocabulary size.

Table 3: Grade one vocabulary size and English scores on the university entrance exam

	n	Mean	Standard Deviation
Vocabulary Size	139	4006	418.79
Entrance Exam Result	139	200.9	19.62

Table 3 shows that 139 first grade students provided their vocabulary size (M = 4006, SD = 418.79) and their university English exam results (M = 200.99, SD = 19.62). A Pearson's r data analysis revealed a correlation, r (137) = .22, p = .008. Thus, it can be assumed that the students with a higher vocabulary generally had a better score on the entrance exam.

Table 4: Grade two vocabulary size and TOEIC scores

	n	Mean	Standard Deviation
Vocabulary Size	74	4002.70	454.04
TOEIC Score	74	612.12	115.85

Table 4 shows that 74 second grade students provided their vocabulary size (M = 4002.70, SD = 454.04) and their most recent TOEIC score (M = 612.12, SD = 115.85). A Pearson's r data analysis revealed a correlation, r (72) = .25, p = .03. The students with a higher vocabulary were likely to have a better TOEIC score.

The study showed that a higher vocabulary correlated with higher scores in both the TOEIC and the university entrance examination. The X-Lex (Meara & Milton, 2003) is an effective way to measure student vocabulary size. It is well suited for educators as it does not take a lot of time to implement: it is also good for researchers because it produces a lot of data in a relatively short amount of time. This is also beneficial for students, because unlike other vocabulary tests it does not rely on skills like reading comprehension, which can lead to fatigue. Despite the benefits mentioned above there are some limitations. As mentioned earlier the test may not be suited for higher intermediate learners as the test does not reflect their knowledge above 5000 words.

Conclusion

The study measured the orthographic vocabulary size and development of students in the first two years of university. The results showed that the vocabulary size of the students was surprisingly high at 4174 when they first started university,

but decreased to 3980 by the start of the second year. This is contrary to the steady increase that previous studies suggest should be expected over time. It was suggested that the decrease is due to vocabulary focus not being as high or important as it is in junior and senior high school. In junior and senior high school having a high vocabulary is necessary to pass exams that directly impact on students' futures. Educators must endeavour to assign more value to vocabulary retention.

The X-Lex (Meara & Milton, 2003) test correlated with both TOEIC and the English entrance exam. It also supported the frequency model. However, for this particular group of participants the knowledge of the first five 1000 word frequency bands was unexpectedly high. As a result, it may underestimate the true vocabulary size of the participants. In future studies it may be beneficial to expand the test to the first eight 1000 word frequency bands.

The results in this study show that educators need to place more value on vocabulary retention. The participants have been extrinsically motivated to study vocabulary for years in junior and senior high school. The participants in this study are non-English majors and therefore may not have the motivation to study and retain vocabulary. All teachers and course books regularly introduce new vocabulary. However, not all of them incentivise vocabulary retention. This can be done by simply creating vocabulary tests and making it part of the course grade. It appears that the Japanese junior and senior high school system produces students with excellent orthographic vocabulary knowledge and everything should be done to make sure that the strong foundations are consistently built upon.

Limitations

This study was completed during the COVID-19 pandemic. As a result all university lessons were moved online. An online version of the vocabulary test was easily adaptable. However, this meant that the researcher was unable to monitor the participants during the test. Participating in the test was optional and it did not impact the student grades in anyway. Therefore, there is no reason to suspect that

the participants did not complete the test as instructed.

References

Aizawa, K. 1998. Developing a vocabulary size test for Japanese EFL learners. *ARELE: Annual Review of English language education in Japan* 9, 75–85.

Aizawa, K. 2006. Rethinking frequency markers for English-Japanese dictionaries. In M. Murata, K. Minamide, Y. Tono & S. Ishikawa (eds), *English Lexicography in Japan* (pp. 108–119). Tokyo: Taishukan-shoten.

Aizawa, K. & Iso, T. 2008. Identifying the minimum vocabulary size for academic reading. *Annual Review of English Language Education in Japan* 19, 121–130.

Anderson, R. C. & Freebody, P. 1981. Vocabulary knowledge. In J. T. Guthrie (ed), *Comprehension and Teaching: Research Reviews* (pp. 77–117). Newark, DE: International Reading Association.

Arnaud, P. J. 1992. Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate component tests. In P. J. L. Arnaud & H. Bejoint (eds), *Vocabulary and applied linguistics* (pp. 133–145). London, England: Macmillan.

Chiang, H. 2018. English vocabulary size as a predictor of TOEIC listening and reading achievement among EFL students in Taiwan. *Theory and Practice in Language Studies* 8 (2), 203–212.

David, A. 2008. Vocabulary breadth in French L2 learners. *Language Learning Journal* 36 (2), 167–180.

Hu, M. & Nation, I. S. P. 2000. Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13 (1), 403–430.

Gairns, R. & Redman, S. 1986. *Working with Words: A Guide to Teaching and Learning Vocabulary*. Cambridge: Cambridge University Press.

Hu, M. & Nation, I. S. P. 2000. Vocabulary density and reading comprehension. *Reading in a Foreign Language* 13 (1), 403–430.

Kanzaki, M. 2015. Comparing TOEIC and vocabulary test scores. In G. Brooks & M. Stoke (eds), *JALT Conference proceedings*. Tokyo: JALT 64.

Koda, K. 1998. The effects of transferred vocabulary knowledge on the development of L2 reading Proficiency. *Foreign Language Annals* 22, 529–540.

Laufer, B. 1989. What percentage of text is essential for comprehension? In C. Lauren & M. Nordman (eds), *Special Language; from Humans Thinking to Thinking Machines* (pp. 316–323). Clevedon: Multilingual Matters.

Laufer, B. 1992. How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Bejoint (eds), *Vocabulary and applied linguistics* (pp. 126–132). London, England: Macmillan.

Meara, P. 1992. *EFL Vocabulary Tests*. University College Swansea: Centre for Applied Language Studies.

Meara, P. & Jones, G. 1990. *Eurocentre's Vocabulary Size Test: User's Guide*. Zurich: Eurocentres.

Meara, P. & Milton, J. 2003. *X-Lex, The Swansea Levels Test*. Newbury: Express.

Meara, P. & Wolter, B. 2004. V-Links, beyond vocabulary depth. *Angles on the English Speaking World World* 4, 85–96.

Milton, J. 2006. X-Lex: The Swansea Vocabulary Levels Test. In C. Coombe, P. Davidson & D. Lloyd (eds), *Proceedings of the 7th and 8th Current Trends in English Language testing (CTELT) Conference*, Vol. 4 (pp. 29–39). UAE: TESOL Arabia.

Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

Milton, J. & Meara, P. 1998. Are the British really bad at learning foreign languages? *Language Learning Journal* 18, 68–76.

Mizumoto, A. & Shimamoto, T. 2008. A Comparison of aural and written vocabulary size of Japanese EFL University Learners. *Language Education & Technology* 45, 35–52.

Mochizuki, M. 1998. A Japanese size test for Japanese learners of English. *IRLT Bulletin* 12, 27–53.

Mochizuki, M. 2010. Development of a vocabulary test battery estimating English skills and proficiency. Report of the Grant-in-Aid for Scientific Research (no. 19320084) by ministry of Education, Science, sports and culture.

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary. Boston*, MA: Heinle and Heinle.

Nation, I. S. P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. 2007. Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton & J. Treffers-Daller (eds), *Modelling and Assessing Vocabulary Knowledge* (pp. 33–43). Cambridge: Cambridge University Press.

Orosz, A. 2009. The growth of young learners' English vocabulary size. In Nikolov, M. (ed), *Early Learning of modern foreign languages*. Bristol: Multilingual Matters, 181–194.

Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Richards, B. J. & Malvern, D. D. 2007. Validity and threats to the validity of vocabulary

measurement. In H. Daller, J. Milton & J. Treffers-Daller (eds), *Modelling and Assessing Vocabulary Knowledge* (pp. 79–92). Cambridge: Cambridge University Press.

Sato, R. 2003. Developing the vocabulary size test for junior high school and high school students According with the new course of study. *STEP BULLETIN* 15, 25–37.

Schmitt, N., Schmitt, D. & Clapham, C. 2001. Developing and exploring the behaviour of two new Versions of the Vocabulary Levels Test. *Language Testing* 18, 55–88.

Seashore, R. H. & Eckerson, L. D. 1940. The measurement of individual differences in general English Vocabulary. *Journal of Educational Psychology* 31, 14–38.

Staehr, L. S. 2008. Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal* 36 (2), 139–152.

Taguchi, K. 2015. An exploratory study of correlations among vocabulary size, vocabulary learning strategy, *TOEIC scores and self-efficacy. Economic Review of Toyo Univesity* 41 (1), 55–67.

Vermeer, A. 2004. The relation between lexical richness and vocabulary size in Dutch L1 and L2 children. In P. Bogaards and B. Laufer (eds), *Vocabulary in a Second Language* (pp. 173–189). Amsterdam: John Benjamins.

Wesche, M. & Paribaht, T. A. 1996. Assessing second language vocabulary knowledge: Depth versus Breadth. *The Canadian Modern Language Review* 53, 13–40.

Wilkins, D. A. 1972. *Linguistics in Language Teaching*. London: Arnold.